

Editorial

Data sharing: Some points of view for scrutiny

Sri Lanka Journal of Child Health, 2017; **46**(2): 105-107

DOI: <http://dx.doi.org/10.4038/sljch.v46i2.8261>

(Key words: Data sharing)

The phenomenon of data sharing has evoked a recent reawakening of interest in the scientific world. This has been particularly so with respect to the medical literature. Data sharing is the implementation of a system of making data obtained and used for learned research, available and accessible to other investigators. From time immemorial, replication and imitation have an extended history in scholastic and academic science. In fact in the present day and age, such ventures are even in line with the motto of The Royal Society, '*Nullius in verba*', which when interpreted means, 'take no man's word for it'¹. In addition, from an ethical perspective, data sharing has assumed an ever so important and extraordinary position in today's academia because transparency, honesty and openness are considered by most discerning individuals to be a part of the scientific method.

The process of sharing data and records created from scientific research has two primary objectives. They can be used to confirm the original analysis and also for formation of hypotheses². From the medical scientific literature perspective, it has the potential to facilitate new discoveries, enhance clinical care of those afflicted by disease and even increase knowledge percolated from data collected in these original scientific works. In fact, data sharing is also likely to be viewed as an ethical and technical necessity. Most certainly, some data sharing procedures have created specific arguments and debate as to which data should be shared, with whom and how quickly³⁻⁶. In those respects however, in spite of several publications, there is only limited reliable information to arrive at a useful and informed decision.

There have been many a discussion around the actual need for sharing of data and these have projected several important key issues. One constructive argument has been that the evidence platform resulting from primary trial publications may not be quite complete and if there is no data sharing, the information from these studies may be lost to research, science and also to patients². The worst scenario is that such data, if they are not shared, would not be available for re-analysis. The classic example of the way in which sharing of data from multiple studies is beneficial is the pivotal

contribution made to science by meta-analysis of published data and the impact of such endeavours in advancing clinical care. In fact, if there is sharing of data and partnership among researchers, there may even be better probabilities for deriving new perceptions from meta-analysis of individual participant data. Routine sharing of clinical trial data could facilitate individual participant data meta-analyses, with the objective of achieving assessments of the effects of interventions that are expected to be more exact, accurate, and unswervingly applicable to diverse types of patients; a component of precision medicine⁷. Needless to say that such meticulous commitment to excellence would only be of benefit to our patients.

It is well known that some investigators may, perhaps even intentionally, withhold and do not report vital findings from their scientific studies. As classic examples, earlier reports involving rofecoxib⁸ and oseltamivir⁹ elucidate the problem of imperfect or deceptive reporting of clinical trial data. The currently employed clinical trial registrations with the exhaustive appraisal of trial registration information, study protocols and statistical analysis plans together with careful scrutiny of manuscripts by editors, peer reviewers and readers, have perhaps reduced erroneous reporting of clinical trials. However, none of these could eradicate the problem totally. By making all information available for re-analysis, data sharing could facilitate research ventures remaining faithful to the trial plans and may go some way towards preventing deliberate withholding of crucial results.

It is now evident that re-analysis may form an important component of ensuring scientific validity of the conclusions and recommendations emanating from scientific inquiry. However, there is a dearth of sharing of data in many research ventures. In one study, Navar *et al* document the practice of data sharing that has been undertaken in three clinical trial databases to which several pharmaceutical companies had contributed data from 3255 trials¹⁰. Through 2015, 154 studies from 177 proposals that had been handled and met original requirements had been approved. The results indicate that although data were available from over 3000 trials through these databases, just about 15.5% had been

sought by a limited number of investigators. This implies noticeably reduced utilisation of clinical trial data, even when they are made freely available, by the researchers. A previous publication had explored Medline from commencement through March 2014 and identified 37 re-analyses that were reported in 36 published articles¹¹. These authors found that 13 of the re-analyses led to changed expositions compared to the original article. However, they did not document whether the re-analyses necessarily changed the inferences of the trial findings with regard to either on-going research or clinical care of patients.

It is imperative that the authorities concerned take all necessary steps to create a data sharing system that protects the privacy and confidentiality of patients, is efficient, effective, and fair to the investigators who have secured the original data. One needs to look at several concerns before implementing a credible system. As an ethical responsibility to the study participants, the shared data must be de-identified for the safety of the subjects. This must be done through a clear and authenticated process. Moreover, during the informed consent procedure, the subjects will need to be made aware that their de-identified data may be shared and provision must be made for them to have the occasion to decline participation because of this explicit requirement. The actual system of sharing data must be effective and robust. Any system of data sharing must also be reasonable, fair and respect the outlay and contributions of the original trial investigators. At the same time, it must be undeniably acknowledged that all these considerations will escalate the complexities of carrying out clinical trials. At a stage of the current academic world where funding for Randomised Clinical Trials (RCTs) is declining, a requirement to share data may also lead to a reduction in the numbers of trials that would be conducted¹². Considering the amount of time and effort invested in conducting RCTs, it has been postulated that the requirement to share the data quickly may become a deterrent towards conducting such trials¹².

In a study published very recently on the use of the National Heart, Lung, and Blood Institute Data Repository, it has been noted that the demand for trial data for secondary analysis has been increasing¹³. However, there is a dearth of information as to what happens once data are released for widespread data sharing. There have also been some concerns about the details and the aftermath of data sharing. It has been argued that someone not involved in the generation and collection of the data may not fully comprehend the choices made in defining the parameters¹⁴. Special difficulties may also arise if data are to be

combined from independent studies and unilaterally considered to be comparable. Questions have been asked as to the comparability of study populations, eligibility criteria, heterogeneity of different study populations and essentials of actual means and devices engaged in different studies¹⁴. Another worrying aspect is the possible advent of a new class of investigators; “research parasites”, consisting of persons who were not involved in the original primary study but who use another group’s data to secure their own ends. Some erudite researchers have even expressed concerns as to whether the system will be taken over by the said research parasites.

There is new impetus in the erudite publishing territory to open up data on the citations that connect research publications. In a unique and novel exercise, 6 Founder Institutions have announced this year the establishment of the Initiative for Open Citations. It has an attractive modern day acronym; I4OC¹⁵. Twenty nine Participating Publishers and 34 Stakeholders have joined this initiative. The main function of I4OC is to formally make openly accessible citations. The resolve of I4OC is to harmonise ventures to encourage the formation of a wide-ranging, freely accessible body of academic citation data. It is the considered opinion of many that this initiative would be of immense value for new as well as existing services and would facilitate the mining, exploration and reuse of the data to add to the current knowledge.

An excerpt from the Editorial published on 21st January 2016 in the *New England Journal of Medicine*¹⁴ and reproduced *verbatim* here, is as follows: “How would data sharing work best? We think it should happen symbiotically, not parasitically. Start with a novel idea, one that is not an obvious extension of the reported work. Second, identify potential collaborators whose collected data may be useful in assessing the hypothesis and propose a collaboration. Third, work together to test the new hypothesis. Fourth, report the new findings with relevant co-authorship to acknowledge both the group that proposed the new idea and the investigative group that accrued the data that allowed it to be tested. What is learned may be beautiful even when seen from close up”. These are words of wisdom from a very reputed journal. That pathway outlined certainly seems to be the way to go. There will always be challenges that would need to be met to make data sharing a worthwhile exercise. In the final reckoning, we also have to recognize the fact that, as far as the medical literature goes, any helpful benefits resulting from sharing of data would only assist the care of the people who ultimately matter; the patients.

References

1. History of the Royal Society. Available from: <https://royalsociety.org/about-us/history/>. Accessed on 18th April 2017
2. Bauchner H, Golub RM, Fontanarosa PB. Data sharing: An ethical and scientific imperative. *Journal of the American Medical Association* 2016; **315**(12):1238-40. <https://doi.org/10.1001/jama.2016.2420> PMID: 27002444
3. Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: the wall must come down. *Journal of the American Medical Association* 2013; **309**(13):1355-6. <https://doi.org/10.1001/jama.2013.1299> PMID: 23508736
4. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. *Journal of the American Medical Association* 2015; **313**(8):793-4. <https://doi.org/10.1001/jama.2015.292> PMID: 25594500
5. Krumholz HM, Peterson ED. Open access to clinical trials data. *Journal of the American Medical Association* 2014; **312**(10):1002-3. <https://doi.org/10.1001/jama.2014.9647> PMID: 25203080
6. Christakis DA, Zimmerman FJ. Rethinking reanalysis. *Journal of the American Medical Association* 2013; **310**(23):2499-500. <https://doi.org/10.1001/jama.2013.281337> PMID: 24346985
7. Berlin JA, Golub RM. Meta-analysis as evidence: building a better pyramid. *Journal of the American Medical Association* 2014; **312**(6):603-5. <https://doi.org/10.1001/jama.2014.8167> PMID: 25117128
8. Karha J, Topol EJ. The sad story of Vioxx, and what we should learn from it. *Cleveland Clinic Journal of Medicine* 2004; **71**(12):933-9. <https://doi.org/10.3949/ccjm.71.12.933> PMID: 15641522
9. Loder E, Tovey D, Godlee F. The Tamiflu trials. *British Medical Journal* 2014; **348**:g2630. <https://doi.org/10.1136/bmj.g2630> PMID: 24811414
10. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. *Journal of the American Medical Association* 2016; **315**(12):1283-4. <https://doi.org/10.1001/jama.2016.2374> PMID: 27002452
11. Ebrahim S, Sohani ZN, Montoya L, et al. Reanalyses of randomized clinical trial data. *Journal of the American Medical Association* 2014; **312**(10):1024-32. <https://doi.org/10.1001/jama.2014.9646> PMID: 25203082
12. Rosenbaum L. Bridging the Data-Sharing Divide - Seeing the Devil in the Details, Not the Other Camp. *New Eng J Med*. Published on-line first 26th April 2017. Available from: <http://www.nejm.org/doi/full/10.1056/NEJMp1704482#t=article>. Accessed 27-04-2017. <https://doi.org/10.1056/NEJMp1704482>
13. Coady SA, Mensah GA, Wagner EL, Goldfarb ME, Hitchcock DM, Giffen CA. Use of the National Heart, Lung, and Blood Institute Data Repository. *New Eng J Med*. Published on-line first 29th March 2017. Available from: <http://www.nejm.org/doi/full/10.1056/NEJMSa1603542#t=article>. Accessed 21-04-2017.
14. Longo DL, Drazen JM. Data Sharing. *New England Journal of Medicine* 2016; **374**:276-7. <https://doi.org/10.1056/NEJMe1516564> PMID: 26789876
15. Initiatives for Open Citations (I4OC) Launches with Early Success. Available from: <http://www.arl.org/news/arl-news/4256-initiative-for-open-citations-i4oc-launches-with-early-success#.WPxAGaJiPIU>. Accessed on 23rd April 2017.

B.J.C. Perera

Joint Editor

 <http://orcid.org/0000-0001-7789-8793>

The author declares that there are no conflicts of interest.

Open Access Article published under the Creative

Commons Attribution CC-BY License. 